LSA.365 | Variation in Optimality Theory, Handout #1
Arto Anttila

VARIATION, TYPOLOGY, UNIVERSALS

1. Phenomena

(1) Variation in form (= variation), variation in meaning (= ambiguity)

Meaning		Meaning$_1$ Meaning$_2$
Form$_1$ Form$_2$		Form

(2) Far from being a marginal "performance" phenomenon, variation is central to human language and an exciting source of evidence about its structure.

(3) Examples of variation in phonology, morphology, and syntax:
- (a) English: *west bank* ~ *wes' bank*
- (b) Finnish: *má.ke.a* ~ *má.kee* ' sweet', *lá.si.-a* ~ *lá.si-i* 'glass-PAR'
- (c) Finnish: *omeno-iden* ~ *omen-ien* ~ *omeno-itten* ~ *omena-in* 'apple-PL.GEN'
- (d) English: *Transactions of the Philological Society* ~ *Philological Society's Transactions* (Jespersen 1949:314)

(4) Variation may have sociolinguistic and stylistic conditions (external factors):
- (a) English *t,d*-deletion [see references in Labov 1997]
- (b) Vowel Coalescence in Colloquial Helsinki Finnish (Paunonen 1995:112, Anttila in press): favored by young working class speakers; disfavored by old upper class speakers, especially females.
- (c) Finnish Genitive Plural: no evidence for sociolinguistic or stylistic conditioning in *omeno-iden* ~ *omen-ien* ~ *omeno-itten*, whereas *omena-in* sounds old-fashioned, e.g. in Mika Waltari's (1908-1979) early novels.
- (d) English possessives: ?

(5) Variation may have grammatical conditions (internal factors, this course):
- (a) English *t,d*-deletion depends on onset structure, neighboring segments (OCP), morphology, and lexical frequency (Bybee 2002, Coetzee 2004, Côté 2000, Guy 1980, 1991ab, 1994, 1997ab, Guy and Boberg 1997, Guy and Boyd 1990, Guy and Myers 1997, Kiparsky 1993, Labov 1997):

MORE DELETION	LESS DELETION	CONDITION
west bank, lost Lenny	*west end, lost Renny*	syllable structure
list (cf. *act*[ə]*d*, Nycz 2005)	*lift*	segmental OCP
los+t	*toss#ed*	level 1 vs. level 2
pass#ed	*kiss#ed*	lexical frequency

(b) Finnish Vowel Coalescence depends on vowel height, morpheme boundary, and lexical category (Anttila in press):

MORE COALESCENCE	LESS (OR NO) COALESCENCE	CONDITION
súo.me-a ‘Finnish-PAR’	*rúot.si-a* ‘Swedish-PAR’	vowel height
lá.si-a ‘glass-PAR’	*rá.si.a* ‘box-PAR’ (**rá.sii*)	morph. bound.
má.ke.a ‘sweet, a.’	*í.de.a* ‘idea, n.’ (**í.dee*)	lex. category

(c) Finnish morphology is sensitive to optional secondary stress which is itself sensitive to syllable weight and vowel height [more later].

(d) English possessives (e.g. Anttila and Fong 2004, Rosenbach 2005):

MORE S-GENITIVE	MORE OF-GENITIVE
animate NP*, Kim*	inanimate NP, *the book*
high theta role, e.g. agent	low theta role, e.g. theme
discourse-old NP	discourse-new NP
non-relational head, e,g, *cat*	relational head, e.g. *picture, performance*
short NP	long NP (Is this phonology?)

Both variation in form and variation in meaning:
their performance vs. *the performance of them*

(6) Research questions:
  (a) The locus of variation: Why does variation occur in certain environments, but is blocked in others?
  (b) The degrees of variation: What determines the quantitative preferences among the variants?
  (c) Grammar and context: How are internal and external factors related?

(7) Variation cannot be reduced to phonetics, sociolinguistics, etc. It is deeply embedded in grammar and refers to grammatical variables:
- onset structure, segmental OCP, stress, syllable weight, vowel height
- level ordering, morpheme boundaries, lexical categories
- animacy, thematic roles, discourse novelty, lexical semantics

2. Modeling variation in OT

2.1 English *t,d*-deletion

(8) *t,d*-deletion in American English:

(a)	It cost ~ cos five dollars.	(*t* before a consonant)
(b)	It cost ~ cos us five dollars.	(*t* before a vowel)
(c)	That’s how much it cost ~ cos.	(*t* before a pause)

(9) Data from five dialects (Coetzee 2004: 218):

		_C	_V	_##
Chicano English (Los Angeles)	*n*	3,693	1,574	1,024
(Santa Ana 1991:76, 1996:66)	% deleted	62	45	37
Tejano English (San Antonio)	*n*	1,738	974	564
(Bayley 1995:310)	% deleted	62	25	46
AAE (Washington, DC)	*n*	143	202	37
(Fasold 1972:76)	% deleted	76	29	73
Jamaican mesolect (Kingston)	*n*	1,252	793	252
(Patrick 1991:181)	% deleted	85	63	71
Trinidadian acrolect	*n*	22	43	16
(Kang 1994:157)	% deleted	81	21	31
Neu data	*n*	814	495	--
(Neu 1980:45)	% deleted	36	16	--

(10) (a) In all dialects, deletion rate is highest in _C.
(b) Deletion rates in _V and _## may occur in either order.

(11) Optimality Theory (Prince and Smolensky 1993/2004):
(a) Grammatical constraints make potentially conflicting structural demands.
(b) Conflicts among constraints are resolved by strict ranking.
(c) Constraints are universal, rankings are language-specific.

(12) Constraints (Kiparsky 1993):

*COMPLEX	Avoid consonant clusters within a syllable.
ONSET	Syllables have onsets.
PARSE	Segments belong to syllables.
ALIGN-LEFT-WORD	Syllables cannot straddle word boundaries.
ALIGN-RIGHT-PHRASE	Phrase-final consonants are also syllable-final.

(13) Sample ranking. Winners: *cost us* (no deletion), *cos me* (deletion), *cos* (deletion)

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost us	(a) [cost][us]	*!	*			
	(b) [cos]t[us]		*!			*
	(c) → [cos][tus]			*		
cost me	(a) [cost][me]	*!				
	(b) → [cos]t[me]					*
	(c) [cos][tme]	*!		*		
cost	(a) [cost]	*!				
	(b) → [cos]t				*	*

2.2 Universals

(14) "So many candidates, so few optima!" (Prince 2006, p. 27)

(15) What kinds of languages are predicted? Use OTSOFT (Hayes, Tesar, and Zuraw 2003) to compute the <u>factorial typology.</u> *t,d*-deletion is highlighted.

```
With 5 constraints, the number of logically possible grammars is 120.
There were 6 different output patterns.

              Output #1    Output #2    Output #3    Output #4
/cost us/:    [cost][us]   [cos]t[us]   [cos]t[us]   [cos][tus]
/cost me/:    [cost][me]   [cos]t[me]   [cos]t[me]   [cost][me]
/cost/:       [cost]       [cost]       [cos]t       [cost]

              Output #5    Output #6
/cost us/:    [cos][tus]   [cos][tus]
/cost me/:    [cos]t[me]   [cos]t[me]
/cost/:       [cost]       [cos]t
```

(16) Only four types of deletion systems are predicted (Kiparsky 1993:4):
- (a) Deletion before C
- (b) Deletion before {C, V}
- (c) Deletion before {C, pause}
- (d) Deletion before {C, V, pause}, i.e. everywhere.

(17) A sample universal:

If <*cost us,* [*cos*]*t*[*us*]> is in a dialect, so is <*cost me,* [*cos*]*t*[*me*]>.

(18) This implicational universal follows from the constraints in (12), but it was not obvious from either the tableau or the factorial typology. Maybe there are more?

(19) There are 5 implicational universals hidden in (14):

```
<cost us, cos]t[us>     -->   <cost me, cos]t[me>
<cost us, cost][us>     -->   <cost me, cost][me>
<cost us, cost][us>     -->   <cost, cost]>
<cost, cos]t>           -->   <cost me, cos]t[me>
<cost me, cost][me>     -->   <cost, cost]>
```

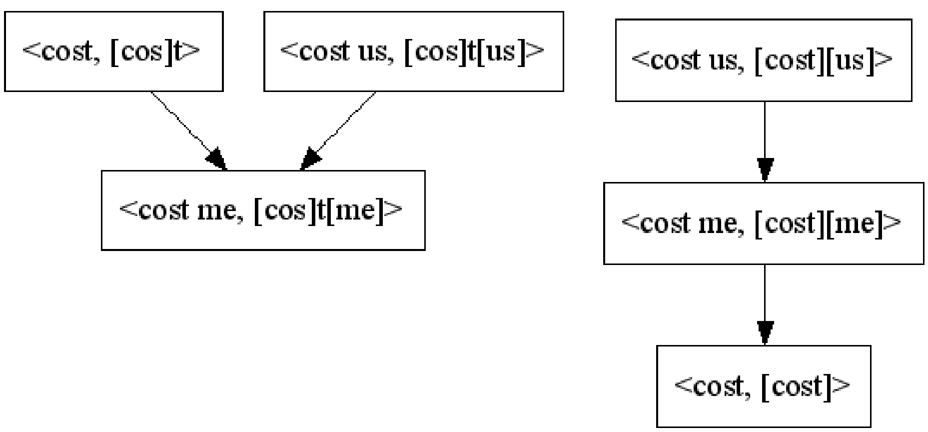
(20) <u>T-order</u>: The set of all implicational universals in a factorial typology

(21) Problem: Factorial typologies are hard for humans to understand. Working out the T-order with paper and pencil is tedious.

(22) Solution (Anttila and Andrus 2006):

> **T-order-Generator is a Windows program that takes a factorial typology as input and returns the corresponding T-order as a directed graph.**

(23) The T-order for the constraint set in (12)



(24) Constructing the T-order:

- For all <input, output> pairs in the factorial typology, construct all the directed edges consisting of a start pair and an end pair, with different inputs.
- For each edge <$pair_0$, $pair_1$>, look through all the output patterns in the factorial typology. If for some output pattern, $pair_0$ appears but $pair_1$ does not, discard the edge. If $pair_1$ appears whenever $pair_0$ appears, keep the edge.

2.3 Quantitative universals

(25) How about variation and quantitative patterns?

(26) Assume the Multiple Grammars Theory (Kiparsky 1993, Anttila to appear):
(a) Variation arises from multiple grammars within/across individuals.
(b) The number of grammars predicting an output is proportional to its frequency of occurrence.

(27) Example: Assume an individual with three total rankings of types {#1, #5, #6}. In the long run, this individual's *t,d*-deletion rates will approximate 0, 2/3, and 1/3.

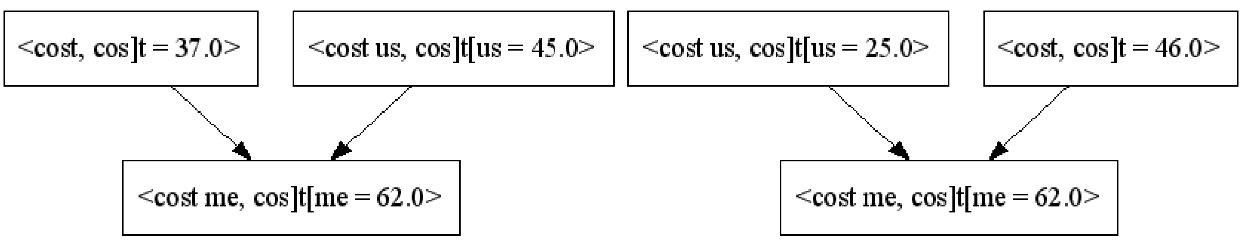
	Output #1	Output #5	Output #6	Del. rate
/cost us/:	[cost][us]	[cos][tus]	[cos][tus]	0/3
/cost me/:	[cost][me]	[cos]t[me]	[cos]t[me]	2/3
/cost/:	[cost]	[cost]	[cos]t	1/3

(28) Consequence: There is no combination of grammars with more *t,d*-deletion before vowels or pauses than before consonants.

(29) In other words, the numbers can only grow in the direction of the arrows.

(30) Quantitative predictions = the empirical generalizations in (10):
- (a) Deletion rate before {V, pause} can never exceed deletion rate before C.
- (b) Nothing is predicted about the ordering of V and pause.

(31) Examples:
- (a) Chicano English (Los Angeles), Santa Ana 1991:76, 1996:66
- (b) Tejano English (San Antonio), Bayley 1995:310



(32) Two types of quantitative patterns:
- (a) Quantitative universals are ranking-independent and follow directly from the constraints (= T-order).
- (b) Quantitative particulars are ranking-dependent and hence subject to cross-linguistic variation.

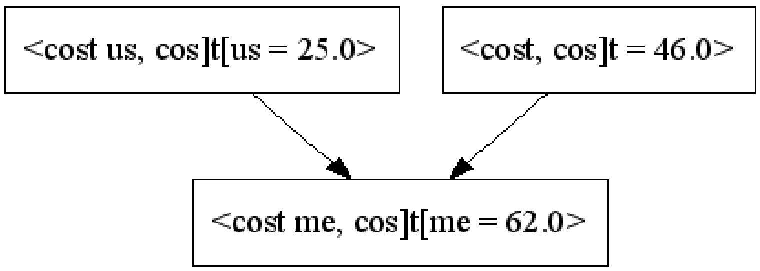
(33) Quiz: A T-order can never contain edges where the start pair has the same input as the end pair, e.g. <cost me, cost][me> --> <cost me, cos]t[me>. Why not? (In fact, you might even try to interpret this edge as saying something quite reasonable: 'Before a consonant, *t,d*-deletion is more common than *t,d*-retention'. What is wrong with this reasoning?)

(34) Question: How well does the T-order fit the data?

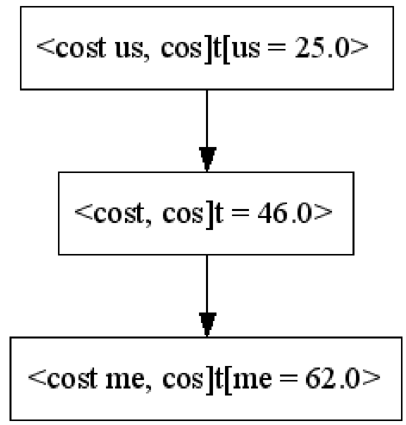
(35) Answer: Check the match between the actual T-order and the ideal T-order:
- (a) *A* = the edges predicted by the grammar
- (b) *I* = the edges that capture all the quantitative relationships in the data

(36) Actual and ideal T-orders for Tejano English as directed graphs

(a) The actual T-order



(b) The ideal T-order



(37) Actual and ideal T-orders for Tejano English as pairs of pairs

```
(a)   <cost us, cos]t[us> --> <cost me, cos]t[me>     actual      ideal
(b)   <cost, cos]t> --> <cost me, cos]t[me>           actual      ideal
(c)   <cost us, cos]t[us> --> <cost, cos]t>                       ideal
```

(38) Two evaluation metrics:
- (a) Precision = $| A \cap I | / | A |$, i.e. how many of the predicted edges are correct
- (b) Recall = $| A \cap I | / | I |$, i.e. how many of the correct edges are predicted

(39) In our example, precision = 1.0, recall = 0.666666666667

(40) A note on the computation of recall:
- A T-order never contains edges with identical start and end pairs (alternative outputs for the same input). For this reason, such edges are not included in the ideal T-order either and thus do not figure in the computation of recall.
- In contrast, edges with exactly the same empirical number form a cycle in the ideal T-order and figure in the computation of recall.

(41) What do these numbers mean?
- (a) Perfect precision (= 100%) indicates DESCRIPTIVE SUCCESS: the universals hold up in one particular language.
- (b) Perfect recall (= 100%) indicates EXPLANATORY SUCCESS: all the quantitative relationships in the data are universal.

(42) A good theory should maximize both precision and recall, i.e. it should be both true and informative.

2.4 The role of constraint ranking

(43) How can adding rankings change the shape of the T-order? Answer: Possibly add arrows into the T-order, but never subtract any existing arrows.

(44) Three types of relations between two <input, output> mappings:

(a) Neither mapping implies the other (no implication)

	#1	#2	#3	#4	#5
_V:	[cost][us]	[cos]t[us]	[cos]t[us]	[cos][tus]	[cos][tus]
_##:	[cost]	[cost]	[cos]t	[cost]	[cos]t

(b) One mapping implies the other (= implication):

	#1	#2	#3	#4
_V:	[cost][us]	[cos]t[us]	[cos][tus]	[cos][tus]
_C:	[cost][me]	[cos]t[me]	[cost][me]	[cos]t[me]

(c) Both mappings imply the other (= equivalence):

	#1	#2	#3
_/a/:	[cost][us]	[cos]t[us]	[cos][tus]
_/i/:	[cost][it]	[cos]t[it]	[cos][tit]

(45) Adding rankings into the grammar will either eliminate columns or have no effect. Clearly, adding rankings can never add new columns into the factorial typology because new rankings restrict the set of languages; they never expand it. Consider the possible effects of ranking in each case:

- In (41a), eliminating #2 would add an arrow, so would eliminating #5.
- In (41b), eliminating #4 would add an arrow, resulting in equivalence.
- In (41c), adding arrows is not possible.

2.5 Summary

(46) Why are T-orders interesting?

(a) T-orders are a consequence of OT, not a new theoretical device.

(b) T-orders are universal and do not have to be learned.

(c) T-orders are general: they hold true under several theories of variation, e.g. Multiple Grammars (Kiparsky 1993), Partially Ordered Grammars (Anttila and Cho 1998), and Stochastic OT (Boersma and Hayes 2001). This is because in all these theories the factorial typology is the same.

(d) T-orders diagnose the adequacy of constraints. If the T-order conflicts with the empirical typology, you will need different constraints.

(e) T-orders express relative well-formedness across inputs [more later].

3. Beyond factorial typologies

(47) Question: Why must we go through the factorial typology? Shouldn't it be possible to compute the T-order directly from the constraint violation tableau?

(48) A failed attempt: Let A, B be two <input, output> pairs. If A incurs a superset of B's violations, then A precedes B in the T-order. Example: <*cost us,* [*cos*]*t us*> violates both ONSET and PARSE, <*cost me,* [*cos*]*t me*> only violates PARSE; and the former precedes the latter in the T-order.

(49) Constraint violations

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost us	(a) [cost][us]	*	*			
	(b) [cos]t[us]		*			*
	(c) [cos][tus]			*		
cost me	(a) [cost][me]	*				
	(b) [cos]t[me]					*
	(c) [cos][tme]	*		*		
cost	(a) [cost]	*				
	(b) [cos]t				*	*

(50) <*cost us,* [*cos*]*t us*> vs. <*cost me,* [*cos*]*t me*>

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost us	(b) [cos]t[us]		*			*
cost me	(b) [cos]t[me]					*

(51) Problem: Consider the exhaustively syllabified pairs: <*cost me*, [*cost*] [*me*]> precedes <*cost*, [*cost*]> in the T-order, yet both only violate *COMPLEX.

(52) <*cost me*, [*cost*] [*me*]> vs. <*cost*, [*cost*]>

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost me	(a) [cost][me]	*				
cost	(a) [cost]	*				

(53) The asymmetry must arise from their respective competitor sets, but exactly how?

(54) Prince (2002, 2006) proposes a way to calculate T-orders directly from constraint violation patterns. This is hard if we use "legacy tableau-ware". We need two new notions: Comparative Tableaux and Elementary Ranking Conditions (ERCs).

(55) Acknowledgements: This approach was also pointed out to me independently by John McCarthy (p.c., 2006) and Marc van Oostendorp (p.c., 2006).

(56) Quick evaluation:

- The Prince method is direct, but has not yet been (completely) implemented.
- The Anttila-Andrus method is indirect, but the algorithm is simple (and efficient) and has been actually implemented.

(57) A comparatively annotated tableau for the “delete-always” grammar

INPUTS	OUTPUTS	*COMPLEX	ALIGN-L-W	ONSET	ALIGN-R-P	PARSE
cost us	(a) → [cos]t[us]			*		*
	(b) [cost][us]	* W		*		L
	(c) [cos][tus]		* W	L		L
cost me	(a) → [cos]t[me]					*
	(b) [cost][me]	* W				L
	(c) [cos][tme]	* W	* W			L
cost	(a) → [cos]t				*	*
	(b) [cost]	* W			L	L

Each loser cell is compared to the corresponding winner cell and annotated as follows:

- If the constraint favors the loser, insert L.
- If the constraint favors the winner, insert W.
- If the constraint is neutral, do nothing.

(58) Comparative tableau

INPUTS	OUTPUTS	*CX	A-L-W	ONSET	A-R-P	PARSE
cost us	(a) [cos]t[us] > [cost][us]	W				L
	(b) [cos]t[us] > [cos][tus]		W	L		L
cost me	(a) [cos]t[me] > [cost][me]	W				L
	(b) [cos]t[me] > [cos][tme]	W	W			L
cost	(a) [cos]t > [cost]	W			L	L

(59) <u>Elementary Ranking Condition (ERC)</u>: Some W dominates all L’s.

(60) <u>ERC vector</u>: A row consisting of W, L, *e* (= blank), e.g. (W, e, e, e, L).

(61) Candidate *c* is optimal iff every ERC [*c* > *x*] holds for all *x* in *c*’s candidate set.

(62) Comparatively annotated tableau for the “aligned retention” grammar

INPUTS	OUTPUTS	*COMPLEX	ALIGN-L-W	ONSET	ALIGN-R-P	PARSE
cost us	(a) → [cost][us]	*		*		
	(b) [cos]t[us]	L		*		* W
	(c) [cos][tus]	L	* W	L		
cost me	(a) → [cost][me]	*				
	(b) [cos]t[me]	L				* W
	(c) [cos][tme]	*	* W			
cost	(a) → [cost]	*				
	(b) [cos]t	L			* W	* W

(63) Comparative tableau

INPUTS	OUTPUTS	*CX	A-L-W	ONSET	A-R-P	PARSE
cost us	(1ab) [cost][us] > [cos]t[us]	L				W
	(1ac) [cost][us] > [cos][tus]	L	W	L		
cost me	(2ab) [cost][me] > [cos]t[me]	L				W
	(2ac) [cost][me] > [cos][tme]		W			
cost	(3ab) [cost] > [cos]t	L			W	W

- An ERC with only Ws (here: [2ac]) is satisfied by any ranking. [This reflects the fact that the loser is harmonically bounded.]
- An ERC with only Ls is not satisfied by any ranking.

(64) How does all this relate to T-orders?
  (a) Each <input, output> pair has an ERC set that guarantees its optimality.
  (b) The edges in the T-order are entailments among ERC sets.

(65) Example: <cost me, [cost][me]> entails <cost, [cost]>

cost me	(2ab) [cost][me] > [cos]t[me]	L				W
cost	(3ab) [cost] > [cos]t	L			W	W

(66) “It should be clear that if one desired winner has a subset of the occasions for victory that another one has, or a superset of the occasions for defeat, then the success of the first will ensure that the second one also succeeds” (Prince 2006, p. 20).

(67) Two rules of inference:
  (a) W-extension: Replace any value in an ERC with a W, and the result is entailed by the original.
  (b) L-retraction: Remove an L from any ERC and the result is entailed by the original.

(68) In (65), [cost][me] entails [cost] by W-extension.

(69) Example: <cost us, [cost][us]> entails <cost me, [cost][me]>

cost us	(1ab) [cost][us] > [cos]t[us]	L				W
	(1ac) [cost][us] > [cos][tus]	L	W	L		
cost me	(2ab) [cost][me] > [cos]t[me]	L				W

(70) Here [cost][us] entails [cost][me] because the ERC set for the second is a subset of the ERC set of the first: (1ab) = (2ab).

(71) For a third rule of inference (Fusion) and an algorithm for computing T-orders using ERCs, see Appendix A.

(72) Harmonic Bounding (Prince 2006, p. 27) is a special case of ERC entailment. Harmonic Bounding is limited to <input, output> pairs within a single input. ERC entailments obtain between <input, output> pairs where the inputs may or may not be the same.

(73) Intuitively, T-orders correspond to ranking complexity: the predecessor contains more ranking information than the successor.

(74) The Complexity Hypothesis: The probability of an <input, output> mapping is inversely correlated with its ranking complexity.

4. Summary

(75) An OT grammar defines a set of implicational universals that hold among <input, output> pairs. We call this structure a T-order.

(76) A T-order is simply a relation (partial order) that holds among <input, output> mappings. The nature of the mapping is immaterial.

(77) Three types of <input, output> mappings related in a T-order:
- (a) Within alternation: the same alternation in different environments, e.g. English prevocalic and prepausal deletion: <c*ost us*, [*cos*]*t* [*us*]> → <c*ost me*, [*cos*]*t* [*me*]>
- (b) Across alternations: different alternations in different environments, e.g. Singapore English prevocalic *p*-copy and prepausal *sp*-metathesis (Anttila, Fong, Benus, and Nycz 2004): <lisp-ing, [lipsping]> → <lisp, [lips]>
- (c) Gradient phonotactics: faithful mappings across different environments, e.g. OCP effects in Arabic: <t-d, [t-d]> → </t-s/, [t-s]>

(78) T-orders are reflected both categorically and quantitatively.

5. Optimality-theoretic approaches to variation

(79) The Multiple Grammars Theory is the generic theory of variation.

5.1 Partially Ordered Grammars

(80) Finnish Vowel Coalescence: *má.ke.a~má.kee* 'sweet', *lá.si.-a~lá.si-i* 'glass-PAR'.

(81) *EA Avoid /ea, oa, öä/ hiatus.
*IA Avoid /ia, ua, yä/ hiatus.
FAITH No coalescence.

(82) Coalescence is more common in mid vowels (*ea*) than high vowels (*ia*) (Paunonen 1995:106-114). Solution: Fixed ranking *EA >> *IA

(83) The constraint violation pattern

/suome-a/ 'Finnish-PAR'	FAITH	*EA	*IA
(a) suomea		*	
(b) suomee	*		
/ruotsi-a/ 'Swedish-PAR'	FAITH	*EA	*IA
(c) ruotsia			*
(d) ruotsii	*		

(84) Assuming the fixed ranking *EA >> *IA, we get the following factorial typology:

		suome-a	*ruotsi-a*
(a)	FAITH >> *EA >> *IA	*suome-a*	*ruotsi-a*
(b)	*EA >> FAITH >> *IA	***suome-e***	*ruotsi-a*
(c)	*EA >> *IA >> FAITH	***suome-e***	***ruotsi-i***

(85) A Partially Ordered Grammar (Anttila 1997, Anttila and Cho 1998, Auger 2001, Ringen and Heinämäki 1999, Zamma 2005; most analyses in Reynolds 1994) is a binary relation (= set of ordered pairs) which is irreflexive, asymmetric, and transitive. Any partial order can be translated into a set of total orders, but not vice versa.

(86) Within the factorial typology in (84) there are 6 partial orders:

(a,b,c)
Colloquial Helsinki Finnish
*EA >> *IA
ea~ee, ia~ii

(a,b)
Old Töölö
*EA >> *IA
FAITH >> *IA
ea~ee, ia

(b,c)
Western Uusimaa
*EA >> *IA
*EA >> FAITH
ee, ia~ii

(a)
Literary Finnish
*EA >> *IA
FAITH >> *EA
FAITH >> *IA
ea, ia

(b)
General Häme
*EA >> *IA
*EA >> FAITH
FAITH >> *IA
ee, ia

(c)
(unattested)
*EA >> *IA
*EA >> FAITH
*IA >> FAITH
ee, ii

(87) The typology of invariant dialects:

(a)	*suome-a*	*ruotsi-a*	No coalescence anywhere.
(b)	***suome-e***	*ruotsi-a*	More coalescence in *ea* than *ia*.
(c)	***suome-e***	***ruotsi-i***	Coalescence everywhere.

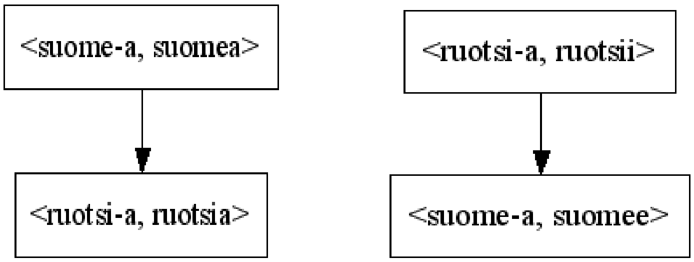
(88) The typology of variable dialects:

(a,b)	*suome-a*	*ruotsi-a*	More coalescence in *ea* than *ia*.
	suome-e	*ruotsi-a*	
(b,c)	***suome-e***	*ruotsi-a*	More coalescence in *ea* than *ia*.
	suome-e	***ruotsi-i***	
(a,b,c)	*suome-a*	*ruotsi-a*	
	suome-e	*ruotsi-a*	More coalescence in *ea* than *ia*.
	suome-e	***ruotsi-i***	

(89) The generic Multiple Grammars Theory allows an additional variable grammar where *ea* and *ia* coalesce at an equal rate:

(a,c)	*suome-a*	*ruotsi-a*	No coalescence anywhere.
	suome-e	***ruotsi-i***	Coalescence everywhere.

(90) Partially Ordered languages are a subset of Multiple Grammars languages, but the T-order remains the same.



5.2 Stochastic Optimality Theory

(91) In StOT (Boersma 1998, Boersma and Hayes 2001, Hayes to appear, Zuraw 2000, among others) each constraint is associated with a real-number <u>ranking value</u>.

(92) <u>Stochastic candidate evaluation</u>: A random positive or negative value ("noise") is temporarily added to the ranking value of each constraint at evaluation time. The resulting <u>selection points</u> are normally distributed around the ranking value.

(93) <u>Gradual Learning Algorithm</u> (GLA, Boersma and Hayes 2001) for StOT:
- (a) Input: a set of arbitrarily ranked constraints, </input/, [output]> pairs
- (b) If the current ranking generates the current learning datum, do nothing. If the current ranking does not generate the current learning datum, then
  - For every constraint violated by the learning datum (= desired winner), decrease its ranking value by a small step.
  - For every constraint violated by the current winner (= wrong winner), increase its ranking value by a small step.

(94) GLA can cope with variation because the adjustments are very small.

(95) Finnish Vowel Coalescence (regular nouns only):

		FREQ	*EA	*IA	FAITH$_{Rt}$	FAITH
(a) /suome-a/	suomea	421	*			
'Finnish-PAR'	suomee	293				*
(b) /idea/	idea	12	*			
'idea'	idee	0			*	*
(c) /ruotsi-a/	ruotsia	4045		*		
'Swedish-PAR'	ruotsii	1014				*
(d) /lattia/	lattia	847		*		
'floor'	lattii	0			*	*

(96) The result of a representative test run (50,000 learning trials, 2,000 test cycles): FAITH$_{root}$ = 106.240, FAITH = 101.306, *EA = 100.568, *IA = 98.126. Average error 2.414 % per candidate

F$_{root}$						F	*EA		*IA
107	106	105	104	103	102	101	100	99	98

(97) Observations and predictions:

	OBS%	PRED%	EXAMPLE	GLOSS	*n*
/e-a/, NOUN	0.410	0.420	/suome-a/	'Finnish-PAR'	714
/ea/, NOUN (recent)	0.000	0.016	/idea/	'idea'	12
/i-a/, NOUN	0.200	0.130	/ruotsi-a/	'Swedish-PAR'	5,059
/ia/, NOUN	0.000	0.001	/lattia/	'floor'	847

(98) Predicts **idee, *lattii* marginally. Increasing the number of exposures will eliminate these forms: with 100,000,000 learning trials and 50,000 test cycles, FAITH$_{root}$ reached the ranking value 114.000, a virtually categorical pattern.

(99) The same factorial typology, hence the same T-order.

(100) If all constraints have the same standard deviation, then stable grammars like {C >> A >> B}, {A >> C >> B}, {A >> B >> C} should not be possible (cf. Multiple Grammars and Partial Ordering).

5.3 The Rank-Ordering Model of EVAL (Coetzee 2004)

(101) Rank-Ordering Model of EVAL (ROE) (Coetzee 2004):
(a) The constraint ranking within a language is a total order.
(b) Candidates that survive past CUT-OFF are all accessible, but rank-ordered.
(c) Several survivors give rise to (i) variation with relative preferences among variants; (ii) gradient word-likeness judgments.

(102) Example: *t,d*-deletion

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost us	(a) [cost][us]	*!	*			
	(b) → [cos]t[us]$_2$		*			*
	(c) → [cos][tus]$_1$			*		
cost me	(a) [cost][me]	*!				
	(b) → [cos]t[me]					*
	(c) [cos][tme]	*!		*		
cost	(a) [cost]	*!				
	(b) → [cos]t				*	*

(103) Input: /cost us/ /cost me/ /cost/

Input:	/cost us/	/cost me/	/cost/	
Outputs:	[cos][tus]$_1$	[cos]t[me]	[cos]t	Decreasing well-formedness
	[cos]t[us]$_2$			

(104) Prediction: No cumulative effects.

(105) This is different in the Multiple Grammars world. Consider a Partially Ordered Grammar *COMPLEX >> the rest: [*cos*]*t*[*us*] (= deletion) wins by 1/3 of the total rankings; [*cos*][*tus*] (= resyllabification) wins by 2/3 of the total rankings.

INPUTS	OUTPUTS	*COMPLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
cost us	(a) [cost][us]	*!	*			
	(b) → [cos]t[us]		*			*
	(c) → [cos][tus]			*		
cost me	(a) [cost][me]	*!				
	(b) → [cos]t[me]					*
	(c) [cos][tme]	*!		*		
cost	(a) [cost]	*!				
	(b) → [cos]t				*	*

(106) An actual example: Finnish allomorph selection (Anttila 1997, 59). Constraints:

*H/I	Avoid heavy syllables with a high vowel nucleus (high = *i, e*).
L.L	No adjacent light (= CV) syllables.
*Í	No stressed high vowels (high = *i, e*).

(107) *náa.pu.rèi.den* ~ *náa.pu.ri.en* 'neighbor-PL-GEN'

/naapuri/ + /-eiden/, /-ien/	*H/I	*L.L	*Í
(a) náa.pu.rèi.den = 37.2%	**		*
(b) náa.pu.ri.en = 62.8%	*	*	

(108) INTRA-CONTEXTUAL and INTER-CONTEXTUAL effects (Coetzee 2004:12-23):

		_C	_V	_##
Jamaican mesolect (Kingston)	*n*	1,252	793	252
(Patrick 1991:181)	% deleted	85	63	71

(a) Deletion is more common than retention <u>within each context</u>.
(b) Deletion is more common <u>before consonants than vowels</u>.

(109) Predicting intra-contextual quantitative effects in ROE (Coetzee 2004, 12-16):

*Ct#C No word-final [t,d] preceded by C and followed by C.
*Ct#V No word-final [t,d] preceded by C and followed by V.
MAX No *t,d*-deletion
Rankings vary across dialects.

(a) Jamaican English

		*Ct#C	*Ct#V	MAX
/Ct#C/	(a) $\varnothing_1$			*
	(b) t_2	*		
		*Ct#C	*Ct#V	MAX
/Ct#V/	(a) $\varnothing_1$			*
	(b) t_2		*	

(b) Tejano English

		*Ct#C	MAX	*Ct#V
/Ct#C/	(a) $\varnothing_1$		*	
	(b) t_2	*		
		*Ct#C	MAX	*Ct#V
/Ct#V/	(a) $\varnothing_2$		*	
	(b) t_1			*

(c) Neu data

		MAX	*Ct#C	*Ct#V
/Ct#C/	(a) $\varnothing_2$	*		
	(b) t_1		*	
		MAX	*Ct#C	*Ct#V
/Ct#V/	(a) $\varnothing_2$	*		
	(b) t_1			*

(110) Multiple Grammars capture intra-contextual effects in exactly the same way: different (sets of) rankings result in quantitative differences in the output.

(111) Predicting inter-contextual quantitative effects in ROE (Coetzee 2004, 16-17):
(a) Assume the fixed ranking: *Ct#C >> *Ct#V.
(b) Assume that EVAL can compare candidates across inputs.

		*Ct#C	*Ct#V	MAX
/Ct#C/	t	*		
		*Ct#C	*Ct#V	MAX
/Ct#V/	t		*	

Since *Ct#C >> *Ct#V, retention in _C is more marked than retention in _V.

(112) Response: Inter-contextual asymmetries follow from the T-order. Neither inherent rankings nor a special evaluation mode is necessary.

Appendix A

A.1 Fusion

(113) A third rule of inference: <u>Fusion</u> (cf. conjunction in propositional logic)

$W \circ W = W$	$W \circ e = e \circ W = W$
$W \circ L = L$	$e \circ L = L \circ e = L$
$L \circ W = L$	$e \circ e = e$
$L \circ L = L$	

This rule is used in the ERC algorithm.

A.2 Computing T-orders using ERCs

(114) The ERC algorithm (Andrus, with input from Anttila and Prince):

```
for each output candidate o:
      -Compute E[o], the set of ERC vectors that must be satisfied to
      ensure that o wins.
      -Throw E[o] out if it contains an invalid ERC vector.

for each ERC set E[o] corresponding to input/output pair (i, o):
      -Look at every ERC set E[p] corresponding to pair (j, p) where i
      and j are distinct inputs.
      -If E[o] entails E[p] then add the edge ((i, o) --> (j, p)) to
      the T-Order.

      This edge belongs in the T-Order because whenever o wins, then
      all the ERCs in E[o] must be satisfied, and by entailment E[p]
      must also be satisfied, so p will also win.

-A set of ERC vectors A entails another set B iff for each vector x in
B, there is some subset Y of A that entails x, i.e. Y fuses to a vector
y such that y entails x.

-Entailment between two ERC vectors x and y can be computed using the
following:
      x entails y iff:
            -y has a W wherever x has a W. (W-Extension)
            -x has an L wherever y has an L. (L-Retraction)
```

A.3 The necessity of fusion

One seems to get pretty far by just W-extension and L-retraction. An earlier version of the algorithm only used these two rules and never failed to find all edges in real-life tests:

```
      -A set of ERC vectors A entails another set B iff for each vector
      x in B, there is some vector y in A such that y entails x.
```

However, as Alan Prince (p.c.) points out, the fusion rule is necessary in cases like the following:

(115) The necessity of fusion

(a) A traditional tableau

INPUTS	OUTPUTS	C1	C2	C3
A	→ a		1	1
	b	1		1
	c		2	
B	→ d			1
	e	1		

(b) A comparatively annotated tableau

INPUTS	OUTPUTS	C1	C2	C3
A	→ a		1	1
	b	1 W	L	1
	c		2 W	L
B	→ d			1
	e	1 W		L

(c) A comparative tableau

INPUTS	COMPARISONS	C1	C2	C3
A	a > b	W	L	
A	a > c		W	L
B	d > e	W		L

A = (W,L,e), (e,W,L). By fusion we get (W,L,L), which entails B = (W,e,L) by L-retraction. But you cannot get to B directly from either individual A-vector by W-extension or L-retraction.